Supporting Information

Elkins *et al.* 10.1073/pnas.0801980105

SI Results

Cultivation and Cell Identification. Members of the *Crenarchaeota* were represented by clone types pOPF_10, pOPF_07, and pOPF_12 (OPF = Obsidian Pool Fermentor), which shared 98% similarity to *Thermosphaera aggregans* strain M11TL^T (1) 98% similarity to *Thermofilum pendens* S10TFL (1) and 98% similarity to an uncultured *Pyrobaculum* sp. HVerd019N (GenBank acc. no. DQ441508), respectively. Two closely related organisms within the *Euryarchaeota* were designated pOPF_1 and pOPF_3 and were 98% and 99% similar, respectively, to an uncultivated member of the Archaeoglobales from Obsidian Pool designated OPPDO15 (2). A single member of the *Korarchaeota*, represented by the 16S rDNA clone pOPF_08, was supported by the enrichment culture. The 16S sequence is 99% similar to both the pJP27 sequence type from Obsidian Pool, Yellowstone National Park (YNP) (3), and the pAB5 sequence from Calcite Springs, YNP (4) (see Fig. S1).

FISH Analysis. Cy3-labeled, oligonucleotide probes (*5′-3′*) KR515R (CCAGCCTTACCCTCCCCT) and KR565R (AGTATGCGTGGGAACCCCTC) allowed visualization of the korarchaeal filamentous cells in single probe hybridizations but provided the best results when applied in combination. A crenarchaeota-specific probe, CREN499R (CCARNCTTGCCCCCCGCT) labeled with Alexa488, failed to hybridize to KR515R/KR565R positive filaments, however, showed bright fluorescence when pure cultures of *Thermofilum* sp. S10TFL were used as a positive control (Fig. S2 *A* and 4*B*). Probes KR515R and KR565R did not hybridize to pure cultures of *Thermofilum* sp. S10TFL. Domain-specific probes ARC915 (5′-GTGCTCCCCCGCCAATTCCT-3′) (5) and EUB338 (5′-GCTGCCTCCCGTAGGAGT-3′) (6) also did not hybridize to the *Korarchaeota* positive filaments (data not shown). Addition of sheared fish sperm DNA to the hybridization buffer reduced background fluorescence. The SDS concentration in the hybridization buffer had to be increased to 0.5% (wt/vol) to allow sufficient probe penetration.

Sequencing and Assembly. Sample sequencing was used to determine which libraries represented the highest number of korarchaeal genomic sequences. The fraction of end reads that aligned with the 84.7-kb korarchaeal genomic contig (Fig. S4) was determined. Libraries BFPP and especially BHXI yielded the highest proportion of sequences that overlapped with the korarchaeal 84.7-kb contig and were selected for further sequencing. A total of 23,000 and 11,520 sequencing reads were produced from the BHXI and BFPP libraries, respectively. The average read length was 800 +/− 113 bp for the small-insert BHXI library and 625 +/− 170 bp for the fosmid-based BFPP library. A high number of reads (2,292) from the BFPP library were of insufficient length and were not used for further analysis. An additional 6,184 sequencing reads from the BFPP library were also determined to be from other Obsidian Pool enrichment culture organisms (based on %GC analysis) and were therefore rejected. Attempts at assembling the genome were made using the 23,000 quality sequencing reads from the BHXI library combined with 3,044 reads from the large-insert library. From the assembly data, it was evident that the original DNA sample contained a dominant genome represented by the pOPF_08 clone type and at least two other genomic variants in low concentration. Because the dominate genome was sufficiently enriched relative to other variants (≈10-fold), genomic heterogeneity did not prevent the assembly of the complete genome of the pOPF_08 organism.

Amino Acid Biosynthesis. Although *Ca.* K. cryptofilum appears to be a proficient peptide degrader, it has an extensive set of amino acid biosynthesis enzymes. The predicted repertoire includes a canonical branched chain amino acid pathway and redundant threonine dehydratase and citramalate synthases to produce 2-oxobutyrate, the precursor for isoleucine biosynthesis. Similar to crenarchaeotes and the *Thermococcales*, *Ca.* K. cryptofilum appears to use the homocitrate pathway and genes of the *lysYZWXJK* operon to produce lysine (7). However, it also encodes *dapA*, *dapE*, *dapC*, and *lysA* homologs from the diaminopimelate pathway of lysine synthesis although *dapB*, *dapD*, and *dapF* genes are missing. Also missing are histidine and tryptophan biosynthesis genes, although the prephenate pathway to phenylalanine and tyrosine is complete. Unlike the known *Crenarchaeota*, the korarchaeote is predicted to use the methanogenic version of chorismate synthesis (8). Common pathways are available to produce arginine, proline, threonine, and serine can be made through the phosphorylating pathway.

Mobile Elements. The *Ca.* K. cryptofilum genome contains six genes (Kcr_0344, Kcr_0272, Kcr_1558, Kcr_0611, Kcr_1387, Kcr_0856) that encode OrfB of the IS605 family of transposases (9), which are widespread throughout the Archaea (10). Kcr_1558 and Kcr_1387 share 87% identity, and both alleles are clustered with predicted ISC1913-like elements similar to those found in *Sulfolobus solfataricus* P2 (11). Kcr_0344 and Kcr_0272 are also nearly identical to one another and are closely related to IS605 sequences found in hyperthermophilic crenarchaeotes as well as transposable elements carried by conjugal plasmids (12).

Tree Compatibility. Analysis of concatenated protein alignments hinges on the assumption that the true phylogenies of all genes involved are the same and the observed differences between trees reconstructed from individual genes represent noise. Thus, ideally, genes must be tested for compatibility of their individual phylogenies before concatenation. Practically, when dealing with large set of genes, such as >30 r-proteins and six RNAP subunits, complete rigorous testing becomes unfeasible. Thus we used the following scheme:

The six RNAP subunit alignments were concatenated, and an ML phylogenetic tree was reconstructed from the concatenated alignment. Each original alignment was tested for compatibility with the topology of this tree using the Approximately Unbiased (AU) test (13) as implemented in the TreeFinder program (14). Three subunits (E′, F, and N) were found to be incompatible with the concatenated tree and, accordingly, were rejected; three (β and two chains of β') passed the test (data not shown). The concatenated alignment of these three subunits was used in all further analyses.

Individual ribosomal proteins are generally too short to be used for a meaningful phylogenetic reconstruction individually, so their alignments were concatenated without detailed testing. We further assessed the compatibility of phylogenetic tree topologies between the concatenated r-proteins and concatenated RNAP subunits. First, *M. kandleri* sequences were removed from both datasets as it has been shown earlier that translational and transcriptional machineries of this archaeon appear to be of different origins (15). Maximum-likelihood

Rib+Rpo:
(((((((Aerpe:0.36773495,Hypbu:0.21297609):0.058441252,Stama:0.27448517):0.066622257,((Sulac:0.15015128,Sulto:0.10456212):0.069624523,Sulso:0.1509329):0.2630078):0.08883196,((Calma:0.3253224,(((Pyrae:0.054856231,Pyris:0.067626872):0.025078936,Pyrca:0.059170292):0.10683526,Thete:0.15772165):0.15116257):0.19498017,Thepe:0.3837574):0.091744444):0.084763622,(Censy:0.76822737,Korar:0.6614682):0.069591385):0.072259578,(((Bacsu:0.33698426,Escco:0.44608078):0.10326424,Deira:0.41947042):1.5273464,Naneq:0.64862503):0.086190568):0.066019763,(((Pyrab:0.012910551,Pyrho:0.022003269):0.01624607,Pyrfu:0.023925165):0.052710879,Theko:0.081936192):0.25734593):0.055934894,((Arcfu:0.34082045,((((((Halma:0.11561483,Natph:0.1182512):0.038811422,Halsp:0.13959571):0.038397559,Halwa:0.14293013):0.40280754,Uncme:0.28868764):0.040151522,((((Metac:0.019253391,Metma:0.02656954):0.019116731,Metba:0.040855086):0.1295261,Metbu:0.18549369):0.1495346,Metsa:0.2967868):0.049075929):0.039992591,(Metcu:0.16004155,(Methu:0.19810155,Metla:0.26673314):0.036847236):0.24328027):0.097104881):0.070349984,(Picto:0.21986978,(Theac:0.070725281,Thevo:0.069452574):0.15861092):0.44427765):0.060677377,((Metja:0.14230929,(MetmC:0.015643543,Metmp:0.014565799):0.25379213):0.19458358,(Metst:0.23903843,Metth:0.1303759):0.23852987):0.041906053);

Rib: (((((((Aerpe:0.31980804,Hypbu:0.20615681):0.055498529,Stama:0.25718752):0.066253714,((Sulac:0.16576925,Sulto:0.11668129):0.077594061,Sulso:0.16282142):0.28501161):0.093247323,((Calma:0.33193714,(((Pyrae:0.057120572,Pyris:0.07077073):0.026296476,Pyrca:0.066530324):0.10905667,Thete:0.17545736):0.1438332):0.18642503,Thepe:0.38173286):0.099866176):0.085274407,(Censy:0.82794032,Korar:0.56696967):0.068304266):0.07088327,Naneq:0.68565168):0.044133173,((Bacsu:0.33527378,Escco:0.45321524):0.098950092,Deira:0.41249964):1.3966304):0.033603488,(((Arcfu:0.37551694,(((((Halma:0.13342765,Natph:0.1326585):0.043048517,Halsp:0.15745755):0.043228326,Halwa:0.15991492):0.42703979,(Metcu:0.18289195,(Methu:0.22837977,Metla:0.29610791):0.043372113):0.21833222):0.035515818,(((((Metac:0.021885312,Metma:0.027189481):0.020552123,Metba:0.047812763):0.13917693,Metbu:0.19809341):0.16309816,Metsa:0.34037388):0.049836532,Uncme:0.34735559):0.03902647):0.10982473):0.060987418,(Picto:0.23763709,(Theac:0.084552
977,Thevo:0.085595389):0.18410383):0.45236938):0.058668401,((Metja:0.14052579,(MetmC:0.015882315,Metmp:0.015780284):0.26479904):0.18936641,(Metst:0.25169702,Metth:0.13607037):0.22555101):0.039398826):0.057778373,(((Pyrab:0.014396699,Pyrho:0.024345608):0.014770393,Pyrfu:0.020655354):0.056742521,Theko:0.0853559):0.24336287);

Rpo: (((((((Aerpe:0.52456895,Hypbu:0.24236839):0.069374313,Stama:0.33298331):0.068457671,((Sulac:0.11654009,Sulto:0.077091543):0.052514386,Sulso:0.12574242):0.21556208):0.07859586,((Calma:0.31876506,(((Pyrae:0.050389173,Pyris:0.062111294):0.022680793,Pyrca:0.042864628):0.10570598,Thete:0.11873635):0.17435016):0.2250776,Thepe:0.40257348):0.072575108):0.077953824,Censy:0.71857094):0.064825184,(((Bacsu:0.34625103,Escco:0.4453692):0.10840656,Deira:0.46139837):2.2483465,(Korar:0.90258868,Naneq:0.75175588):0.10023721):0.060014185):0.072094977,(((Pyrab:0.0096322551,Pyrho:0.016736461):0.019982278,Pyrfu:0.033581242):0.043467586,Theko:0.076615985):0.3032898):0.054780751,((Arcfu:0.27501647,(((Halma:0.076647508,(Halsp:0.099864933,Halwa:0.12959176):0.030925598):0.02240101,Natph:0.067741209):0.36352804,(((((Metac:0.013346087,Metma:0.025591969):0.015592053,Metba:0.024615296):0.10901293,Metbu:0.16095282):0.12200913,Metsa:0.20025675):0.038264588,(((Metcu:0.12050679,Methu:0.13685102):0.032523153,M
etla:0.18647727):0.26028804,Uncme:0.21766027):0.038322895):0.047884765):0.073732157):0.094247225,(Picto:0.18714592,(Theac:0.039550499,Thevo:0.033687756):0.10319597):0.43834352):0.063219904,((Metja:0.15386209,(MetmC:0.016300163,Metmp:0.010996696):0.23172211):0.21688718,(Metst:0.21362919,Metth:0.12149914):0.28171906):0.050458485);

phylogenetic trees were reconstructed for the concatenated r-proteins, concatenated RNAP subunits, and the two protein sets concatenated together. All three reconstructions placed *Ca.* K. cryptofilum, *C. symbiosum*, and *N. equitans* into slightly different deep positions in the archaeal tree. Each of the three reconstructed phylogenies was reduced to an "essential constraint" tree, which preserves the deep branching order of *Ca.* K. cryptofilum, *C. symbiosum*, *N. equitans*, *Crenarchaeota* clade, the *Euryarchaeota* clade, and the Bacteria clade but reduces the rest of the tree to a polytomy. Multifurcations in each of the three constraint trees were resolved separately for the concatenated r-proteins and concatenated RNAP subunits. Each of the two protein sets was tested for compatibility with two alternative topologies against its native topology using the AU test.

The tree topologies of concatenated r-proteins and concatenated RNAP subunits were mutually incompatible (AU P value is <0.0001 for both cross comparisons). However, each topology was compatible with the topology of the joint tree that was obtained from the concatenation of both alignments (AU *P* value of 0.44 for the r-proteins and 0.09 for the RNAP subunits). This concatenation provides the "middle ground" topology which is compatible with both original components.

For compatibility testing, the trees are given in the Newick format or can be represented as rooted or unrooted trees using the TreeView program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html)

SI Materials and Methods

Sample Collection and Cultivation. A variety of sediment and water samples were collected from Obsidian Pool, YNP, and ranged from 78°C to 92°C and had a pH of ≈6.5. Similar samples collected near the same time our samples were taken (within days) were subjected to a detailed geochemical analysis by Schock *et al.* (16). The enrichment culture was established by adding several reduced and nonreduced samples (≈400 ml) of sediment and spring water to a double-walled glass vessel with a total volume of ≈800 ml. The Inoculation was performed inside a Coy anaerobic tent (Coy Laboratories) to avoid exposure to air. Heated glycerol was pumped through the outer chamber of the vessel to maintain a constant temperature of 85°C. A gas mixture consisting of N_2/CO_2 (80:20) was constantly bubbled through the culture at a rate of 20 ml/min. The anaerobic growth medium consisted of a modified Allen's formulation (see below) (17) and was continuously added to the vessel at a rate of 8.0 ml/h using a Gilson Minipuls2 peristaltic pump (Gilson). Culture effluent was allowed to overflow through a side port in the cultivation vessel and into a sterile, anaerobic 2 l glass media bottle.

Obsidian Pool fermentor (OPF) medium (modified from M. B. Allen, 1959):

$(NH_4)_2SO_4$ 1.30 g
KH_2PO_4 0.28 g
$MgSO_4$ x 7 H_2O 0.25 g
$CaSO_4$ x 2 H_2O 0.17 g
$CaCl_2$ x 2 H_2O 0.07 g

$FeCl_3$ x 6 H_2O 0.02 g
Na_2SO_4 0.07 g
KNO_3 0.10 g
Peptone 0.50 g
Yeast Extract 0.10 g
$Na_2S_2O_3$ 0.78 g
Na_2S x 6 H_2O 0.13 g
$MnCl_2$ x 4 H_2O (10 mg/ml) 180 μl
$Na_2B_4O_7$ x 10 H_2O (25 mg/ml) 180 μl
$ZnSO_4$ x 7 H_2O (10 mg/ml) 22 μl
$CuCl_2$ x 2 H_2O (10 mg/ml) 5 μl
Na_2MoO_4 x 2 H_2O (1 mg/ml) 30 μl
$VOSO_4$ x 5 H_2O (1 mg/ml) 30 μl
$CoSO_4$ x 7 H_2O (1 mg/ml) 10 μl.
$LiCl/Na_2WO_4/NaSeO_3/Ni(NH_4)_2(SO_4)$ (1 mg/ml each) 10 μl
Wolfe's Vitamin Solution (1,000×) 1.0 ml
Distilled H_2O 1000 ml

Fluorescent *in Situ* Hybridization Analysis. Fresh samples were harvested directly from the Obsidian Pool enrichment culture and washed three times with 1 ml of PBS (pH 7.2). The cells were then spotted onto precleaned gelatin-coated microscope slides in 8-μl aliquots and dried by placing the slides onto a 50°C heat block. Ten microliters of fixative solution [3% paraformaldehyde (wt/vol) in PBS (pH 7.2)] was placed over the dried cells, and the samples were then incubated either at room temperature for 2 h or overnight at 4°C. The fixed cells were washed thoroughly by rinsing the slides with PBS (pH 7.2) followed by an ethanol dehydration series by submerging the slides into a 50%, 70%, and 100% ethanol (vol/vol) bath for 3 min each. The slides were then dried by applying a gentle stream of compressed air. A hybridization chamber was prepared by saturating a folded paper towel with hybridization buffer and placing the towel in a 50-ml conical centrifuge tube. A volume of 8 μl of hybridization solution containing 0.9 M NaCl, 0.01–0.5% (wt/vol) SDS, 100 μg/ml sheared herring sperm DNA, 0.02 M Tris·HCl (pH 7.2), and 0–40% formamide (vol/vol) was applied to each cell spot and the slides were placed in the hybridization chamber and incubated at 46°C for 15 min. After the prehybridization step, 100 ng of the appropriate fluorescently labeled oligonucleotide probe was added to the hybridization buffer and the slides were incubated at 46°C for 3 h. Excess probe was washed from the slide by gently pipetting 2 ml prewarmed wash buffer containing 0.9–0.06 M NaCl (depending of formamide concentration in hybridization buffer), 0.1% (wt/vol) SDS, and 0.02 M Tris·Hcl (pH 7.2) over the cells. The entire slide was then immersed in prewarmed wash buffer and placed in a 48°C water bath for 15 min. The slide was then rinsed with ≈300 ml of deionized H_2O and dried by applying a gentle stream of compressed air. A drop of Prolong Gold mounting solution containing DAPI (Molecular Probes) was applied to each cell spot followed by a 25 × 25-mm coverslip. All microscopic analysis was performed with a Zeiss Axioplan 2 microscope with 100-W mercury arc and halogen illumination sources. Phase-contrast images were collected using either ×63 or ×100 Plan Neofluar oil-immersion objectives with 1.4 and 1.3 numerical apertures, respectively. Images were collected using a Zeiss Axiocam digital camera and processed with Zeiss Axiovision software.

Electron Microscopy. *Korarchaeota* cell pellets were fixed in a solution containing 2.5% glutaraldehyde (EM grade) in 20 mM sodium cacodylate buffer (pH 6.5). For scanning electron microscopy drops of the fixed sample were placed onto glass slides, covered with a coverslip and rapidly frozen with liquid nitrogen. The coverslip was removed with a razor blade, and the slide was immediately transferred into fixative buffer, postfixed with osmium tetroxide, dehydrated in a graded series of acetone solutions and critical-point dried from liquid CO_2, mounted on stubs, and coated with 3 nm of platinum with a magnetron sputter coater. The specimens were examined with a Hitachi S-4100 field emission scanning electron microscope. For negative staining, a drop of the sample at appropriate dilution was placed on a 400 mesh carbon-coated copper grid, freshly hydrophilisized by glow discharge. After incubation for 2 min, the drop was quickly removed with a pasteur pipette and the grid was air dried. The grid was stained with 2% uranium acetate and 0.01% glucose. Micrographs were taken with an EM 912 electron microscope (Zeiss) equipped with an integrated OMEGA energy filter operated in the zero loss mode.

Cell Purification. Culture effluent from the Obsidian Pool fermentor was collected anaerobically in sterile 2-liter glass bottles. The cells were pelleted by centrifugation at 6,000 rpm in a Beckman JLA-10.500 rotor (Beckman Coulter, Fullerton) for 20 min at 25°C in 500-ml centrifuge bottles (Nalge). The cells were resuspended in a total volume of 200 ml of PBS (pH 7.2) and then divided into 4 × 50-ml aliquots in 50-ml conical centrifuge tubes (BD Falcon). The cells were centrifuged again at 6,000 rpm for 20 min in a Beckman JA-12 rotor at 25°C and the supernatant was discarded. Each pellet was resuspended again in 50 ml of PBS (pH 7.2) by vortexing. After the cell pellets were completely resuspended, 0.5 ml of a 20% (wt/vol) SDS solution [final concentration 0.2% SDS (wt/vol)] was added to each tube and immediately mixed by inverting the tube several times. The tubes were centrifuged again at 6,000 rpm for 20 min in a Beckman JA-12 rotor at 25°C. Each pellet was then washed three times by resuspending in 50 ml of PBS (pH 7.2) and centrifuging as described above. After the final wash, the cell suspensions were heated to 85°C and then filtered through 0.45-μm syringe filters (MILLEX HV, Millipore) in 25-ml aliquots. The filtrate was collected in fresh conical centrifuge tubes which were then centrifuged at 6,000 rpm for 30 min in a Beckman JA-12 rotor at 4°C. Small white pellets were resuspended in a total volume of 1 ml of PBS (pH 7.2), placed in an Eppendorf tube, and centrifuged at 16,000 × *g* in a benchtop microcentrifuge. The single-cell pellet was resuspended in either PBS (pH 7.2) for microscopic analysis or DNA isolation.

Genome Sequencing and Assembly. Library construction, high-throughput sequencing, and genome assembly were performed at the Joint Genome Institute, Walnut Creek, CA. Fosmid libraries were constructed from the raw OPF1 enrichment culture DNA and DNA from filtered cells using the pCC1FOS based system (Epicentre Biotechnologies). Small-insert libraries (2–3 kb insert size) were produced for raw, filtered, and purified cell samples of *Ca.* K. cryptofilum. DNA sequencing was performed with BigDye Terminators v3.1 and resolved with ABI PRISM 3730 DNA sequencers (PE Applied Biosystems). The sequencing reads were base called using phred version 0.990722.g (18, 19), vector trimmed using crossmatch SPS-3.57, and assembled into contigs using parallel phrap (www.phrap.org). The contigs were aligned with an 84.7 kb *Ca.* K. cryptofilum genomic sequence containing the pOPF_08 16S rRNA gene, and the 16S rRNA genes from *Thermofilum pendens* and *Thermosphaera aggregans*. The resulting assembly was binned by scaffolding and read depth, and by aligning binned scaffolds against the 16S rRNA sequences from the predominant archaeal genomes in the mixture and an 84.7-kb reference contig.

Gene Prediction and Comparative Analysis. An automated annotation was provided by the Computational Biology Group at Oak Ridge National Laboratory (ORNL). An additional analysis was performed by loading the complete korarchaeal genome into the Integrated Microbial Genomes (IMG) analysis tool developed by the Joint Genome Institute (20). To identify putative functions for the predicted ORFs, the gene models were compared

against all nonredundant (NR) sequences deposited in the National Center for Biotechnology Information (NCBI) database using BLASTP (21). The gene models were also compared against the protein families (Pfam) (22), Clusters of Orthologous Groups (COGs) (23), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (24) databases for further functional information. The BLAST E-value threshold was set at 1e-05. The Pfam threshold was also set at 1e-05. The COG search was performed using the default cutoff. For tRNA identification, tRNAscan-SE was run using the prokaryotic default settings (25), followed by manual analyses. DNA and amino acid sequence alignments were generated with CLUSTALW using the default parameters (26) and checked and edited manually using BioEdit v7.0.5 (www.mbio.ncsu.edu/BioEdit/bioedit.html).

Phylogenetic Analyses. An alignment of concatenated small and large subunit rRNA sequences (SSU+LSU rRNA) was constructed based on their conserved secondary structures and refined by hand using BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html). Highly variable loop regions that could not be confidently aligned as well as regions in which some sequences were incomplete were masked out. The final alignment contained 27 taxa and 2965 nucleotide characters and was analyzed using the software Modeltest v3.7 to determine the best evolutionary model that fits the data (27). The likelihood ratio test statistics indicated that to correspond to the general time reversible model with unequal nucleotide frequencies, a fraction of invariable sites and with six independent substitution rates that follow a gamma distribution (GTR+I+G). Phylogenetic analysis was performed using PAUP* 4.0b8 (Sinauer Associates). The initial GTR+I+G parameters were determined using five rounds of neighbor-joining tree construction/parameter estimation and were used for an unrestricted heuristic maximum likelihood tree search with random sequence addition and tree bisection-reconnection (TBR) branch swapping. The GTR+I+G parameters were estimated from the resulting tree and used for a new round of heuristic search that used five independent rounds of random sequence additions and TBR branch swapping, resulting in a single maximum-likelihood tree. Bootstrapping (100 replicates) was conducted using the same parameters and one round or random sequence addition/TBR branch swapping for each replicate.

Protein sequences were aligned on the M-Coffee web server using a combination of multiple sequence alignment methods (28). The alignments were visually inspected and the regions not confidently aligned or containing high variability were masked out. The final alignments contained 476 amino acid positions for EF2 and 211 amino acid positions for FtsZ genes, respectively. Rooted (Fig. 2*B*) and unrooted (Fig. S6) EF2 trees were both calculated. Phylogenetic reconstruction was performed using two different approaches. First, the amino acid substitution model was chosen with the software Modelgenerator v84, which uses Akaike and Bayesian information criteria (AIC and BIC) to select the optimum among 88 different models (29). For all of the datasets the optimum model was a combination of the RtRev substitution matrix with an estimated fraction of invariable sites and with six categories of substitution rates following a gamma distribution with the shape parameter optimized based on the data. These parameters were used to calculate maximum likelihood trees and non-parametric bootstrap support for the individual nodes with the PhyML v2.4.4 software (30). Since accurate reconstruction of deep phylogenetic branching order is complicated due to long branch attraction artifacts, we also used tree inference with PhyloBayes v2.3 (31). That software uses a Bayesian Monte Carlo Markov Chain (MCMC) sampler and a site-heterogeneous mixture model (CAT) and has been shown to be less prone to LBA (31, 32). Two to four independent chains were run in parallel for at least 10,000 generations starting from random trees, until convergence (maximum discrepancy across all bipartions of <0.05). The first 5,000 cycles were then discarded as the burn-in and the remaining ones were sampled each of every ten to calculate majority-rule posterior consensus trees with bipartition frequencies at each node.

Concatenated r-protein and RNAP trees: Using the COGNITOR method (33, 34), representatives of three bacterial species (*E. coli*, *B. subtilis*, and *D. radiodurans*) were added to the arCOGs that include 33 universally conserved ribosomal proteins and three largest RNA polymerase subunits. The sequences were aligned using MUSCLE (35). Start codon positions were verified and, when necessary, corrected manually. Maximum-likelihood trees for the concatenated alignment of 36 proteins (8,057 amino acid positions, 5,696 for ribosomal proteins, and 2,361 for RNA polymerase subunits) were constructed using TreeFinder (14), with the Whelan and Goldman (WAG) evolutionary model and gamma-distributed site rates. Tree topologies were compared using the Approximately Unbiased test (AU) (13) implemented in the TreeFinder program. The topology where *Ca.* K. cryptofilum is a sister group of all *Crenarchaeota*, including *Cenarchaeum symbiosum* passed the 0.05 *P* value cutoff (AU P value $= 0.17$).

1. Burggraf S, Huber H, Stetter KO (1997) Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data. *Int J Syst Bacteriol* 47:657–660.
2. Spear JR, Walker JJ, McCollom TM, Pace NR (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci USA* 102:2555–2560.
3. Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* 91:1609–1613.
4. Reysenbach AL, Ehringer M, Hershberger K (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and ''Korarchaeota'' coexist. *Extremophiles* 4:61–67.
5. Raskin L, Stromley JM, Rittmann BE, Stahl DA (1994) Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl Environ Microbiol* 60:1232–1240.
6. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56:1919–1925.
7. Nishida H, Nishiyama M, Kobashi N, Kosuge T, Hoshino T, Yamane H (1999) A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Res* 9:1175–1183.
8. White RH (2004) L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose 1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. *Biochemistry* 43:7618–7627.
9. Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62:725–774.
10. Filee J, Siguier P, Chandler M (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* 71:121–157.
11. She Q, *et al.* (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* 98:7835–7840.
12. Greve B, Jensen S, Brugger K, Zillig W, Garrett RA (2004) Genomic comparison of archaeal conjugative plasmids from Sulfolobus. *Archaea* 1:231–239.
13. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
14. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
15. Brochier C, Forterre P, Gribaldo S (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: Tackling the *Methanopyrus kandleri* paradox. *Genome Biol* 5:R17.
16. Shock EL, Holland M, Meyer-Dombard D, Amend JP (2005) in *Geothermal Biology and Geochemstry in Yellowstone National Park*, ed. Inskeep WP (Thermal Biology Institute, Montana State University, Bozeman, MT), pp 95–112.
17. Allen MB (1959) Studies with *Cyanidium caldarium*, an anomalously pigmented chlorophyte. *Arch Mikrobiol* 32:270–277.
18. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194.
19. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
20. Markowitz VM, *et al.* (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34:D344–348.
21. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
22. Finn RD, *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34:D247–D251.

23. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36.
24. Kanehisa M, *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354–357.
25. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
26. Chenna R, *et al.* (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.
27. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
28. Moretti S, *et al.* (2007) The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res* 35:W645–W648.
29. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29.
30. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
31. Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol 7 Suppl* 1:S4.
32. Philippe H, Brinkmann H, Martinez P, Riutort M, Baguna J (2007) Acoel flatworms are not platyhelminthes: evidence from phylogenomics. *PLoS ONE* 2:e717.
33. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637.
34. Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol Direct* 2:33.
35. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.

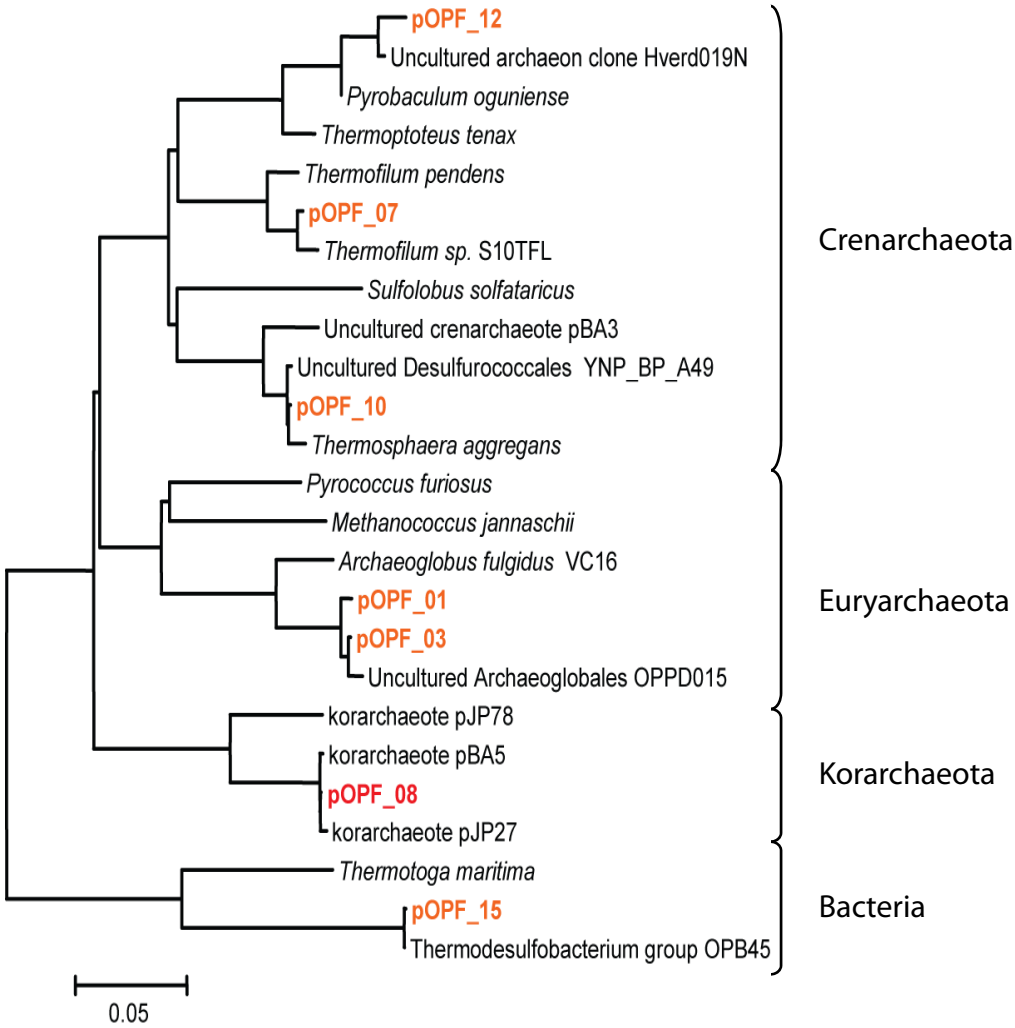

Fig. S1. Neighbor-joining tree constructed from distance matrix analysis of SSU rDNA sequences obtained from the Obsidian Pool enrichment culture. *Ca.* K. cryptofilum is represented by clone pOPF_08.

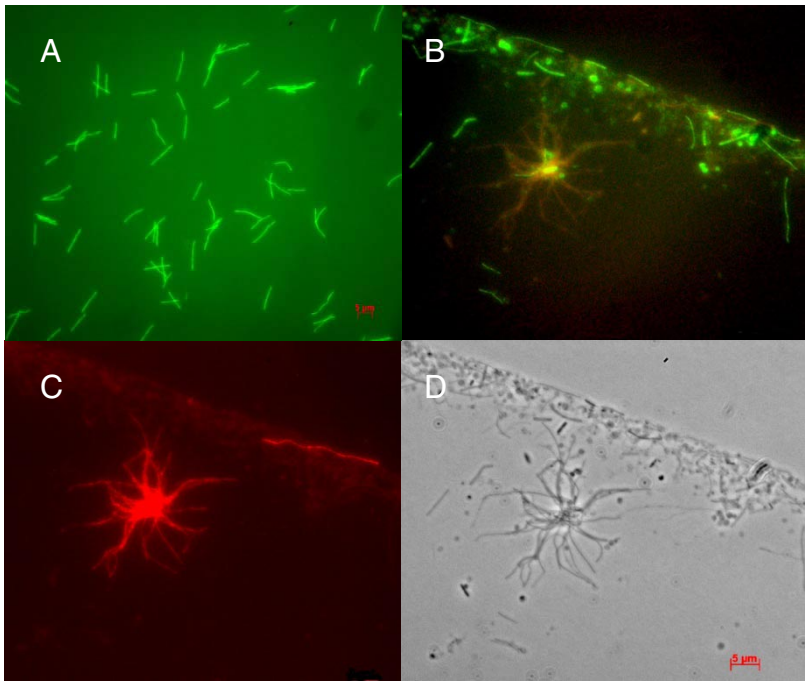

Fig. S2. Fluorescence *in situ* hybridization analysis using crenarchaeal and *Korarchaeota* specific probes. (*A*) Pure culture of *Thermofilum* sp. S10TFL hybridizing to Alexa488-labeled CREN499R (CCARNCTTGCCCCCCGCT) probe. (*B*) CREN499R probe applied to Obsidian Pool enrichment culture. (*C*) Cy3-labeled *Korarchaeota* specific probes KR515R (CCAGCCTTACCCTCCCCT) and KR565R (AGTATGCGTGGGAACCCCTC) hybridizing to cells of *Ca.* K. cryptofilum. (*D*) Phase contrast image of cell preparation. (Scale bars, 5 μm.)

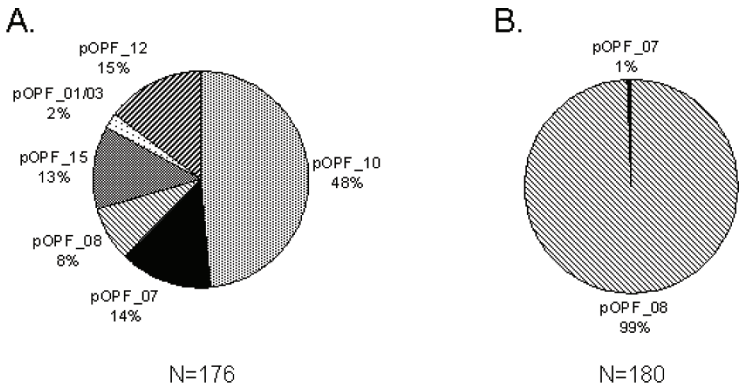

Fig. S3. Relative abundance of PCR amplified SSU rDNAs obtained from (*A*), untreated, Obsidian Pool enrichment culture and (*B*), filtered cells exposed to 0.2% SDS for 15 min. PCR primers included Ab779F (GCRAASSGGATTAGATACCC) [Brunk CF, Eis N (1998) Quantitative measure of small-subunit rRNA gene sequences of the kingdom korarchaeota. *Appl Environ Microbiol* 64:5064–5066] and UA1406R (ACGGGCGGTGWGTRCAA) [Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55:541–555]. The treatment resulted in a highly enriched cell fraction represented by the pOPF_08 SSU rDNA sequence type.

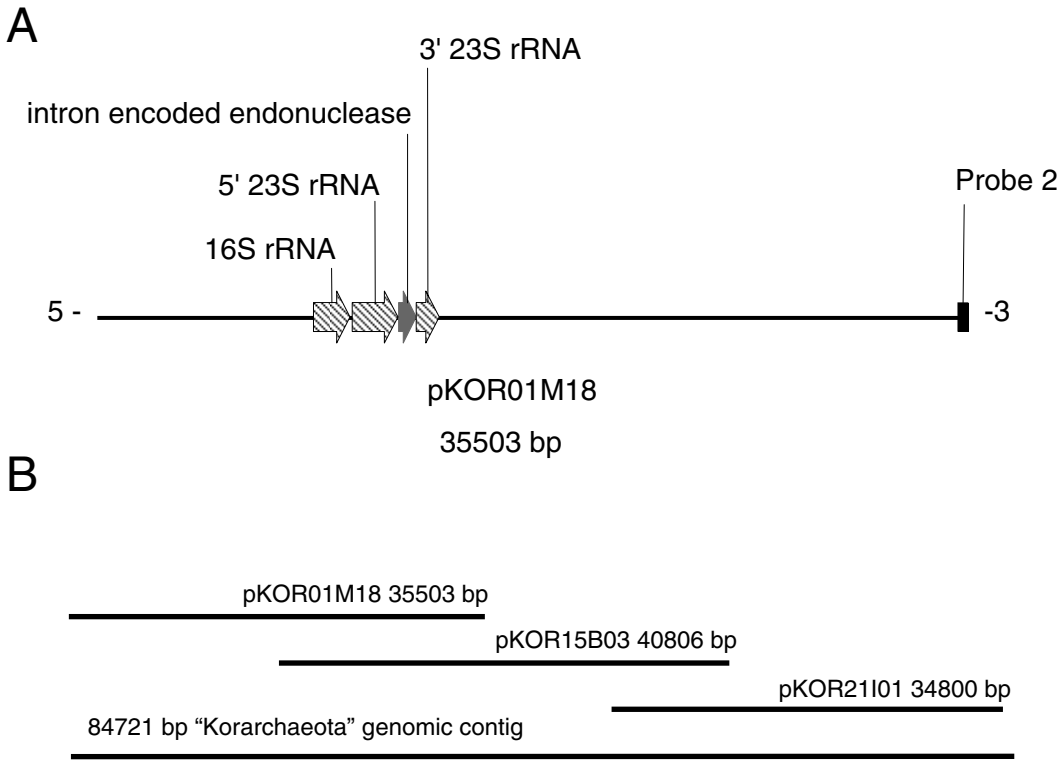

Fig. S4. Before WGS sequencing, fosmid clones containing the pOPF_08 SSU rRNA sequence were retrieved and assembled from a library constructed from Obsidian Pool enrichment culture DNA with a CopyControl fosmid library kit (Epicentre, Madison, WI, USA). (*A*) Fosmid clone pKOR01M18 containing korarchaeal small and large subunit rRNA genes was retrieved by probing the library with a 209 bp, 5′ region of the pOPF_08 sequence. The LSU rRNA gene contains a self splicing, intron- encoded LAGLIDADG type endonuclease. (*B*) Additional clones were retrieved by probing the library with labeled end-sequences from the finished fosmids. Overlapping fosmid clones were assembled to construct an 84.7-kb genomic contig.

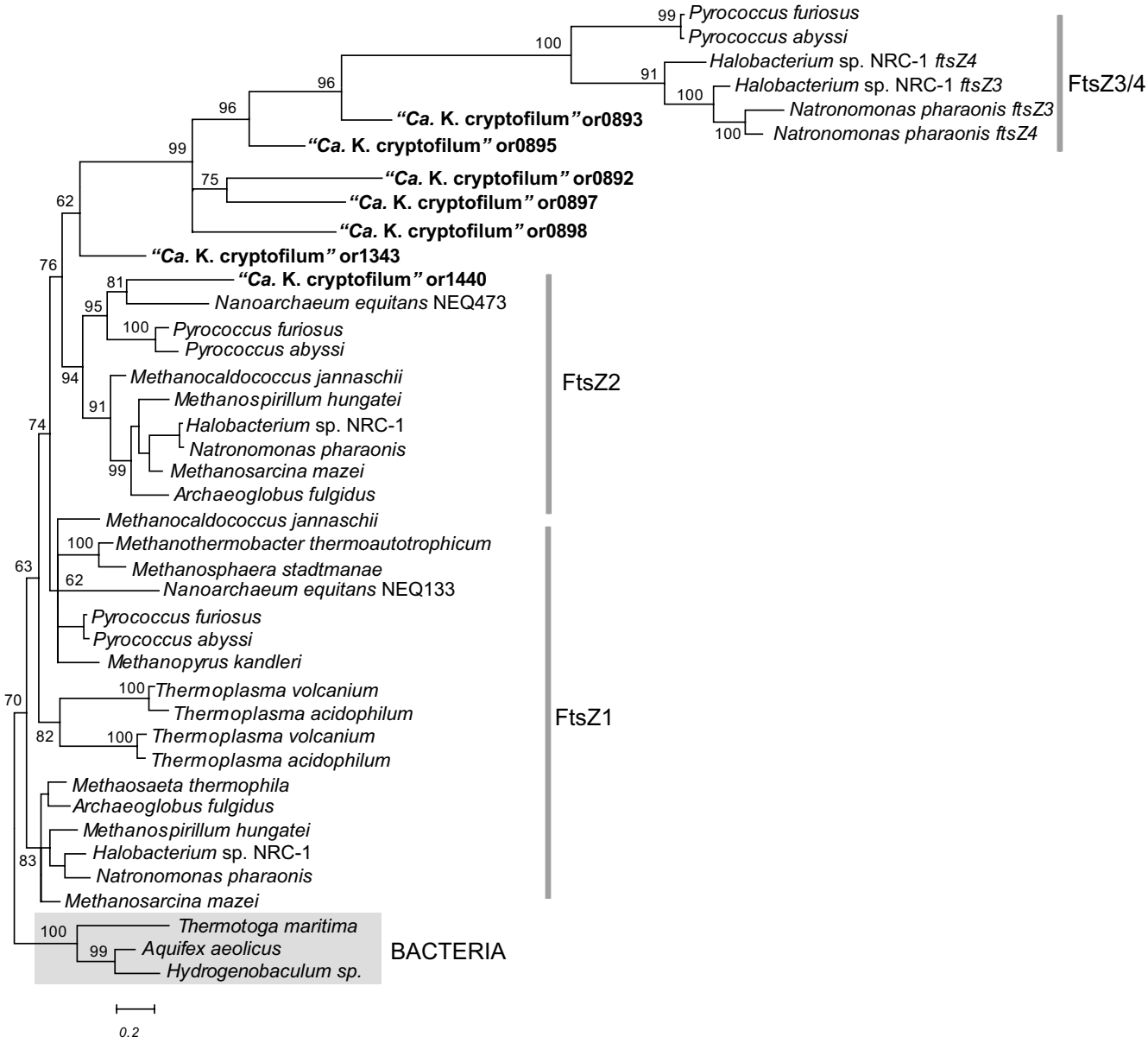

Fig. S5. Majority-rule posterior consensus tree (PhyloBayes) for FtsZ protein sequences. Values at the nodes indicate the posterior probability for that branching (the branch was collapsed when the value was <50). Scale bar indicates the inferred number of substitutions per site.

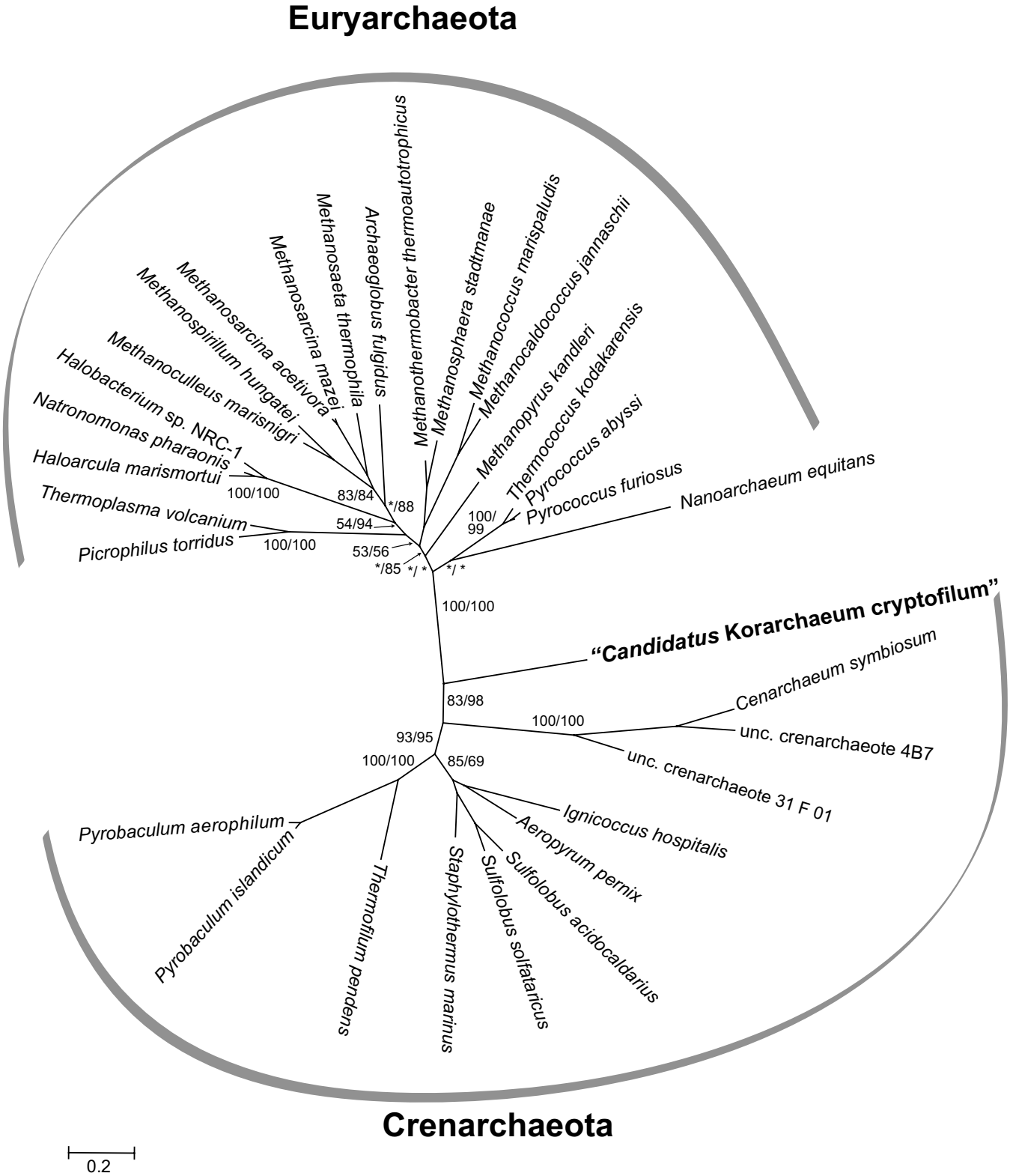

Fig. S6. Unrooted archaeal phylogeny based on translation elongation factor 2 (EF2) proteins. The numbers indicate bootstrap support for PhyML/consensus posterior probability (Phyloblast), an asterisk indicates <50 support. Where both values were <50, the branch was collapsed. Scale bar indicates the inferred number of substitutions per site.

Table S1. Representation of genes that are characteristic of *Crenarchaeota* or *Euryarchaeota* in *Candidatus* K. cryptofilum

Locus tag	arCOG	Functional class	Annotation	No. *Crenarchaeota* genomes	No. *Euryarchaeota* genomes
Genes characteristic of *Euryarchaeota**					
Kcr_1527	**arCOG04447**	**Replication/ repair**	**Archaeal DNA polymerase II, large subunit**	**0**	**27**
Kcr_1385	**arCOG04455**	**Replication/ repair**	**Archaeal DNA polymerase II, small subunit**	**0**	**26**
Kcr_1284	**arCOG00872**	**Replication/ repair**	**ERCC4-like helicase**	**0**	**26**
Kcr_0242	**arCOG02610**	**Replication/ repair**	**Rec8/ScpA/Scc1-like protein (kleisin family)**	**0**	**24**
Kcr_1446	**arCOG02258**	**Replication/ repair**	**RPA family protein, a subunit of RPA complex**	**0**	**20**
Kcr_0892, 0893, 0895, 0897, 0898, 1343, 1440	**arCOG02201**	**Cell division**	**Cell division GTPase FtsZ**	**0**	**26**
Kcr_0243	**arCOG00371**	**Cell division**	**Chromosome segregation ATPase, Smc**	**0**	**24**
Kcr_0227	arCOG01684	transcription	Predicted transcriptional regulator	0	22
Kcr_1298	arCOG03363	Energy metabolism	Archaeal/vacuolar-type H+-ATPase subunit H	0	26
Kcr_0321, 1119	arCOG01549	Energy metabolism	Coenzyme F420-reducing hydrogenase, alpha subunit	0	18
Kcr_1264	arCOG00959	Energy metabolism	Ferredoxin	0	18
Kcr_1262	arCOG01607	Energy metabolism	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit	0	19
Kcr_1100	arCOG04353	Amino acid metabolism	3-dehydroquinate synthase	0	20
Kcr_0835	arCOG02714	Coenzyme metabolism	GTP and metal dependent enzyme involved F420 coenzyme biosynthesis	0	20
Kcr_0481	arCOG00043	General functional prediction only	ATP-utilizing enzyme of the PP-loop superfamily	0	18
Kcr_1356	arCOG00354	General functional prediction only	GTPase SAR1 or related small G protein	0	23
Kcr_0479	arCOG02465	General functional prediction only	NCAIR mutase (PurE)-related protein	0	18
Kcr_0896, 1002	arCOG02155	General functional prediction only	Protein implicated in RNA metabolism, contains PRC-barrel domain	0	26
Kcr_1574	arCOG02673	Uncharacterized	Predicted membrane protein	0	27
Kcr_1412	arCOG02263	Uncharacterized	Uncharacterized conserved protein	0	24
Kcr_1416	arCOG02197	Uncharacterized	Uncharacterized conserved protein	0	23
Kcr_1046	arCOG02408	Uncharacterized	Uncharacterized conserved protein	0	20
Kcr_0480	arCOG02701	Uncharacterized	Uncharacterized conserved protein	0	18
Missing in *Ca.* K. cryptofilum: 47 euryarchaeal-specific genes					
Genes characteristic of *Crenarchaeota*†					
Kcr_0816	**arCOG01013**	**Translation**	**Protein containing a domain similar to ribosomal protein L13E**	**11**	**0**
Kcr_0040	**arCOG04327**	**Translation**	**Ribosomal protein S25**	**13**	**0**
Kcr_1513	**arCOG04293**	**Translation**	**Ribosomal protein S30**	**13**	**0**
Kcr_1196	**arCOG04305**	**Translation**	**Ribosomal protein S26**	**13**	**0**
Kcr_1004	**arCOG04271**	**Transcription**	**DNA-directed RNA polymerase, subunit RPB8**	**12**	**0**
Kcr_0112	**arCOG00393**	**Transcription**	**Predicted membrane-associated transcriptional regulator**	**9**	**0**
Kcr_0586	arCOG00287	Replication/ repair	Predicted NurA-like nuclease	10	0
Kcr_0394	arCOG02299	Secretion/ motility	Peptidase A24A, prepilin type IV	12	0

Locus tag	arCOG	Functional class	Annotation	No. *Crenarchaeota* genomes	No. *Euryarchaeota* genomes
Kcr_0105	arCOG02960	General functional prediction only	Predicted aminopeptidase, Iap family	10	0
Kcr_1526	arCOG04136	General functional prediction only	Uncharacterized Zn ribbon-containing protein	12	0
Kcr_0281, 0620	arCOG01618	General functional prediction only	Aldo/keto reductase, related to diketogulonate reductase	9	0
Kcr_0852	arCOG03721	General functional prediction only	HEPN domain	10	0
Kcr_0392, 0733, 0851	arCOG03722	General functional prediction only	HEPN domain	9	0
Kcr_0387	arCOG03119	Uncharacterized	DedA family membrane protein	13	0
Kcr_1152	arCOG03770	Uncharacterized	Uncharacterized conserved protein	11	0
Missing in *Ca.* K. cryptofilum: 30 crenarchaeal-specific genes					

*Bold type shows genes for replication/repair and cell division system components.
†Bold type shows genes for translation and transcription system components.

Table S2. Bacterial COGs without archaeal members except *Ca*. K. cryptofilum

Locus tag	COG	Function	Top BLASTP Hit	GenBank accession no.	E-value
Kcr_0172	COG0386	Glutathione peroxidase	*Desulfuromonas acetoxidans* DSM 684	ZP_01311479	7.0e-06
Kcr_0823	COG0861	Membrane protein TerC possibly involved in tellurium resistance	*Saccharopolyspora erythraea* NRRL 2338	YP_001105852	2.0e-59
Kcr_1396	COG1349	Transcriptional regulators of sugar metabolism	*Anaeromyxobacter dehalogenans* 2CP-1	ZP_02325805	6.7e-02
Kcr_0831	COG1760	L-serine deaminase	*Algoriphagus* sp. PR1	ZP_01720767	8.0e-70
Kcr_1493	COG1995	Pyridoxal phosphate biosynthesis protein	*Acidiphilium cryptum* JF-5	YP_001234273	1.0e-73
Kcr_1494	COG3395	Uncharacterized protein conserved in bacteria	*Alkaliphilus metalliredigens* QYMF	YP_001318120	5.0e-24
Kcr_1364	COG3485	Protocatechuate 3, 4-dioxygenase beta subunit	*Desulfococcus oleovorans* Hxd3	YP_001530052	5.0e-06

Table S3. Top BLASTP hits for predicted euryarchaeal specific genes in *Ca*. K. cryptofilum

Locus tag	Function	Top BLASTP Hit	GenBank accession no.	E-value
Kcr_0047	Histones H3 and H4	*Methanobacterium formicicum*	P48783	9.0e-14
Kcr_1137	Histones H3 and H4	*Thermococcus zilligii*	P95669	6.0e-14
Kcr_1527	Archaeal DNA polymerase II, large subunit	*Methanosaeta thermophila* PT	YP_843669	0.0
Kcr_1385	Archaeal DNA polymerase II, small subunit	*Methanosaeta thermophila* PT	YP_843089	1.0e-73
Kcr_1440	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_184684	3.0e-68
Kcr_1343	Cell division GTPase	*Methanothermobacter thermautotrophicus* str. Delta H	NP_276787	1.0e-89
Kcr_0892	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_184684	2.0e-38
Kcr_0893	Cell division GTPase	*Pyrococcus horikoshii* OT3	NP_142027	1.0e-28
Kcr_0895	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_183834	2.0e-41
Kcr_0897	Cell division GTPase	*Pyrococcus abyssi* GE5	NP_126968	7.0e-29
Kcr_0898	Cell division GTPase	*Pyrococcus furiosus* DSM 3638	NP_578254	3.0e-42

Other Supporting Information Files

Dataset S1 (XLS)